

REMARKS

Pursuant to the requirements of 37 C.F.R. §§ 1.821-1.825, Applicant submits the enclosed Sequence Listing and computer readable form (CRF). The amino acid sequences disclosed in the specification and drawings may be found in computer readable form in file 010261.txt on the enclosed diskette and are presented in the paper copy of the Sequence Listing, enclosed.

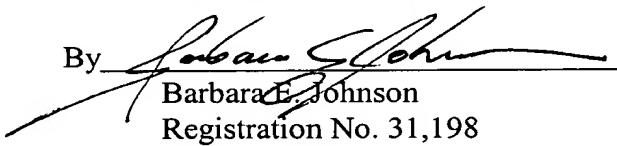
Applicant hereby certifies that the information recorded in computer readable form (CRF) supplied on the enclosed diskette as file 010261.txt is identical to the written Sequence Listing. The material presented in computer readable form is not new matter because it presents sequences the same as those disclosed in the specification, as filed.

Applicant believes that the requirements of 37 C.F.R. §§ 1.821-1.825 have been met.

Respectfully submitted,

WEBB ZIESENHEIM LOGSDON
ORKIN & HANSON, P.C.

By



Barbara E. Johnson
Registration No. 31,198
Attorney for Applicant
700 Koppers Building
436 Seventh Avenue
Pittsburgh, PA 15219-1818
Telephone: 412-471-8815
Facsimile: 412-471-4094



MARKED-UP AMENDED SPECIFICATION PARAGRAPHS

[0018] If the nucleotide sequence is random, the probability that a sequence of given length translated from it will have a particular amino acid sequence can be calculated simply by multiplying together the frequencies in the genetic code of the codons encoding each amino acid [amino acid] in the sequence. Since some amino acids have as many as six codons and others as few as one, the predicted frequency will vary depending on the amino acid sequence itself. Thus the sequence LRRLLR (SEQ ID NO: 1), made up entirely of six-codon amino acids, will appear at a frequency of 1 in $(6/61)^6$, or approximately once in a million codons, and the sequence MWWMMW (SEQ ID NO: 2), made up entirely of one-codon amino acids, will appear at a frequency of 1 in $(1/61)^6$, or approximately once in fifty billion codons. The frequencies of other sequences will fall between these two extremes. The important point for us is that even a relatively short sequence will appear very rarely, and so if we can determine the amino acid sequence of a peptide translated from unknown sequence, we can match it to a portion of the reference sequence with high specificity.

[0030] Comparison of the experimental results with the values in the table indicates reveals a match to the predicted mass value for one of the ten candidates – specifically the sequence that begins at position 3190 of the reference sequence and proceeds from right to left. Retrieval of the reference sequence beginning at position 3190 indicates that the cloned sequence begins with "GAATTCTTACACCTCATACTTCCCAAGCCCCAACTTCTCATCT GAAAATGGTAATAGTATCATCCTTACATGTTAAGGTCATGAATTGCTAT GTGTA.....(1st 100 nucleotides shown) (SEQ ID NO: 3). The identification is confirmed by dideoxy sequencing from a primer 150 nucleotides upstream of the junction between the pUC19 sequence and the EcoRI fragment.

[0032] The peptide TMITPSLHACRSTLED (SEQ ID NO: 4), representing the N-terminal 16 amino acids of the alpha-complementing factor of beta-galactosidase encoded in pUC19 (and also representing the 16 constant N-terminal amino acids in all of the peptides described in Example 1 above) is used to raise a polyclonal rabbit antibody using standard procedures.

[0034] The mass spectrum of the immunoprecipitate from the induced cell lysate of the clone under examination is observed to contain a distinct peak, at a position corresponding to a mass of 8485 ± 3 Daltons, that is not observed in the control. Comparison of the experimental results with the values in the table in example 1 above indicates that the insert begins at position 9241 of the reference sequence and proceeds from left to right in the Genbank sequence. Retrieval of the reference sequence beginning at position 9241 indicates that the cloned sequence begins with

GAATTCACATAAATCGCAAATTTTTTCTCCTCCCAGAGCC
ATCCAAAACTCTGTTGTCAAAGGCCTGTCTGAGGATACCACTGAAGAGA
CATTAAAG.....(1st 100 nucleotides shown) (SEQ ID NO: 5). The identification is confirmed by dideoxy sequencing as described in Example 1.

[0037] To identify the nucleotide sequence adjacent to the pTriplEx' vector, each EcoRI site in the JO5584 sequence is identified and ligated, in silico, to the EcoRI site in the pTriplEx' vector. For each such in silico construct, the amino acid sequences of the two expected hybrid translation products (from each of the start codons in the vector to the first in frame stop codons encountered in the insert) are calculated. The mass of each peptide is calculated and all 10 peptide pairs are tabulated, as shown in the table below. Comparison of the experimental results (i.e., peptides of 4255 and 2635 Da.) with the values predicted in the table indicates that the insert begins at position 4028 of the reference sequence and proceeds in the forward direction. It is concluded that the 5' end of the sequence joined to the vector is

GAATTCTCTGGGTT TTGTGGTGTGCTAGACTTAATTACCCATGAATGATT
TGTCTCTTGAGAAAATTCAATAGCACATCTATTAGTGTAAAA....(1st 100

nucleotides shown) (SEQ ID NO: 6). The identification is confirmed by dideoxy sequencing from the plasmid using a primer 150 nucleotides 3' to the pTriplex' EcoRI site.

<u>Position of EcoRI site</u>	<u>Orientation in pTriplex'</u>	<u>Start Codon</u>	<u>Predicted Peptide Mass</u>
3190	forward	1st	6137
3190	forward	2nd	5707
3190	reverse	1st	6278
3190	reverse	2nd	3891
4208	forward	1st	4255
4208	forward	2nd	2635
4208	reverse	1st	19748
4208	reverse	2nd	3905
6066	forward	1st	3595
6066	forward	2nd	3606
6066	reverse	1st	6401
6066	reverse	2nd	1363
9241	forward	1st	3583
9241	forward	2nd	7122
9241	reverse	1st	4582
9241	reverse	2nd	1746
9543	forward	1st	5306
9543	forward	2nd	1477
9543	reverse	1st	9906
9543	reverse	2nd	2516

The mass values above are computed by translating each hypothetical fusion polypeptide without the N-terminal methionine that is removed in vivo in *E. coli*.

[0040] Two oligonucleotide primers are synthesized using standard methods. In one, CCCGAATTCAGCAGGTAAAAATCAAGG (SEQ ID NO: 7), the first 10 nucleotides contain an EcoRI site (underlined) and last 17 nucleotides correspond to the first 17 nucleotides of exon 2 of the human nucleolin gene. The other, GGGGAATTCTTACTCTTCTCCACTGCTAT (SEQ ID NO: 8), the last 17 nucleotides correspond to the reverse complement of the last 17 nucleotides of exon 2, followed immediately (in the sense orientation of the oligonucleotide) by the stop codon TAA and a sequence that includes an EcoRI site (underlined).

[0044] The program was run with the 24 nucleotide input sequence CAACTAGAAGAGGTAAAGAAACTAT (SEQ ID NO: 9). Two reading frames were selected; the forward reading frame beginning with the first nucleotide (F1) and the reverse (antisense) reading frame beginning with the second antisense nucleotide (R2). The results are shown below.

[begin]

Enter Sequence:

[input] CAACTAGAAGAGGTAAAGAAACTAT (SEQ ID NO: 9)

[output] Protein: QLEEVRYN (SEQ ID NO: 10)

Which reading frames would you like to examine?

1: Forward (F1)

2: Forward; first base removed (F2)

3: Forward; second base removed (F2)

4: Reverse (R1)

5: Reverse first base removed (R2)

6: Reverse second removed (R3)

[input] 1,5

[output] **MASS DIFFERENCES**

<u>Location</u>	<u>Mutation</u>	<u>Frame F1</u>	<u>Frame R2</u>
-----------------	-----------------	-----------------	-----------------

	None	1032.13	722.89
	/A(K)	0.04	0.00
1	C-{ G(E)	0.99	0.00
	\T(Z)	-1032.13	0.00
	/G(R)	28.06	0.00
2	(Q) A-{ T(L)	-14.97	0.00
	\C(P)	-31.01	0.00
	/G(Q)	0.00	0.00
3	A-{ T(H)	9.01	0.00
	\C(H)	9.01	0.00
	/A(I)	0.00	276.34
4	C-{ G(V)	-14.03	276.34
	\T(L)	0.00	0.00
	/C(P)	-16.04	299.37
5	(L) T-{ A(Q)	14.97	226.32
	\G(R)	43.03	200.24
	/G(L)	0.00	241.29
6	A-{ T(L)	0.00	241.33
	\C(L)	0.00	242.28
	/T(Z)	-790.84	-34.02
7	G-{ C(Q)	-0.99	-34.02
	\A(K)	-0.95	0.00
	/G(G)	-72.07	-60.10
8	(E) A-{ T(V)	-29.99	16.00
	\C(A)	-58.04	-44.04
	/G(E)	0.00	-34.02
9	A-{ T(D)	-14.03	-34.02
	\C(D)	-14.03	-48.05
	/T(Z)	-661.72	0.00

10	G-{ C(Q)	-0.99	0.00
	\A(K)	-0.95	0.00
	/G(G)	-72.07	-16.04
11	(E) A-{ T(V)	-29.99	23.98
	\C(A)	-58.04	43.03
	/T(D)	-14.03	0.00
12	G-{ C(D)	-14.03	-14.03
	\A(E)	0.00	34.02
	/T(L)	14.03	-423.52
13	G-{ C(L)	14.03	-423.52
	\A(I)	14.03	0.00
	/C(A)	-28.05	-60.04
14	(V) T-{ A(E)	29.99	-16.00
	\G(G)	-42.08	-76.10
	/G(V)	0.00	-26.04
15	A-{ T(V)	0.00	-49.08
	\C(V)	0.00	-48.09
	/G(G)	-99.14	0.00
16	A-{ T(Z)	-433.47	0.00
	\C(R)	0.00	0.00
	/T(I)	-43.03	76.10
17	(R) G-{ C(T)	-55.09	16.06
	\A(K)	-28.02	60.10
	/G(R)	0.00	10.04
18	A-{ T(S)	-69.11	14.02
	\C(S)	-69.11	-16.00
	/G(D)	0.99	0.00
19	A-{ T(Y)	49.08	0.00
	\C(H)	23.04	0.00

	/G(S)	-27.02	-28.05
20	(N) A-{ T(I)	-0.94	15.96
	\C(T)	-13.00	-42.08
	/A(K)	14.07	48.05
21	C-{ G(K)	14.07	14.03
	\T(N)	0.00	14.03
	/C(H)	-26.04	18.03
	\G(D)	-49.08	0.00
22	T-{ A(N)	-48.09	0.00
	/G(C)	-60.04	-12.06
23	(Y) A-{ T(F)	-16.00	15.01
	\C(S)	-76.10	43.03
	/C(Y)	0.00	-14.03
24	T-{ A(Z)	-163.18	0.00
	\G(Z)	-163.18	0.00

Enter the detection threshold:

[input] 0.8 Dalton.

[output] Undetectable amino acid substitutions: 1.(Q)C-A(K)

[0048] The sequence of exon 2 of the human rds/peripherin gene (Genbank accession M73531) is shown below. Intron sequence is shown in lower case; exon sequence in upper case.

gggaagccatctccagctgtctgttcccttaagTCGAATCAAGAGCAACGTGGATGGCG
 GTACCTGGTGGACGGCGTCCCTTCAGCTGCTGCAATCCTAGCTGCCACGGCCCTGC
 ATCCAGTATCAGATCACCAACAACTCAGCACACTACAGTTACGACCACAGACGGAG
 GAGCTAACCTGTGGGTGCGTGGCTGCAGGGCTGCCCTGCTGAGCTACTACAGCAGCC
 TCATGAACTCCATGGGTGTCGTACGCTCCTCATTGGCTTCAGgtaggccctggcagctg
 gggtagagggtaaggagagcctcc (SEQ ID NO: 11)

[0049]

Two primers, of sequences

GGCCCAGGAATTCTCCAGCTGTCTGTTCCCTTAAG (SEQ ID NO: 12) and AATTTACTCGAGCTACCCCCAGCTGCCAGGGCCTAC (SEQ ID NO: 13) were synthesized and used to PCR amplify rds/peripherin exon 2 from an individual known to carry a wild type allele of rds/peripherin. The amplicon was cut with EcoRI and XhoI and cloned into the EcoRI/XhoI sites of the pGEX derivative described in Nelson et al. The resulting plasmid was cut with Xho 1, treated with Klenow fragment of DNA polymerase, and self-ligated to produce a construct expected to produce a fusion protein with the sequence shown below.

MSPILGYWKIKGLVQPTRLLEYLEEKYEEHLYERDEGDKWRNKKFELGLE
FPNLPYYIDGDVKLTQSMAIIRYIADKHNLGGCPKERAESMLEGAVLDIYGVSIAYSK
DFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLD
AFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLIEGRGIQDLVPH
TTPHHTTPHHTTPHHTPQDLNSPAVCFPLSRIKSNDGRYLVDGVPFSCCNPSSPRPCIQY
QITNNSAHYSYDHQTEELNLWVRGCRAALLSYYSSLMNSMGVVTLIWLFEVPGQLGV
ARSSGRIVTD (SEQ ID NO: 14)

[0053] The amplicons described in the previous example are reamplified using the upstream primer

5'GGATCCTAATACGACTCACTATAGGGAGACCACCATGCATCACCATCACCATCA
CCACTCTCCAGCTGTCTGTTCCCTTAAG (SEQ ID NO: 15) and the downstream primer
5' CTTAGTCATTATACCCCCAGCTGCCAGGGCCTAC (SEQ ID NO: 16). The upstream primer contains a T7 promoter followed by a translation initiation sequence (start codon underlined) followed by a sequence encoding eight histidines followed by sequence identical to the rds/peripherin sequence immediately 5' to rds/peripherin exon 2. The downstream primer

contains two stop codons (in antisense orientation) preceding the sequence complementary to the sequence just 3' to red/peripherin exon 2.

[0061] Because the primers are all anchored by non-T nucleotides at their 3' ends, only three of them will prime a given cDNA sequence. In the case of the hemoglobin alpha 2 transcript, which ends in the sequence

GCGGCAAAAAAAAAAAAAA..., (SEQ ID NO: 17) the primers that are extended are those ending in G.